2011 Seventh IEEE International Conference on eScience

Preparing DARIAH

Tobias Blanke
King's College London
Centre for e-Research (CeRch)
tobias.blanke@kcl.ac.uk

Michael Bryant
King's College London
Centre for e-Research (CeRch)
michael.bryant@kcl.ac.uk

Mark Hedges
King's College London
Centre for e-Research (CeRch)
mark.hedges@kcl.ac.uk

Andreas Aschenbrenner
Austrian Academy of Sciences
Vienna, Austria
andreas.aschenbrenner@oeaw.ac.at

Michael Priddy
Data Archiving and Networked Services (DANS)
The Hague, Netherlands
mike.priddy@dans.knaw.nl

Abstract

This paper analyses the results of the technical and scientific work in the DARIAH preparatory phase, a European infrastructure for digital arts and humanities. We were looking for an infrastructure model that would allow for the integration of services built around communities. To this end, DARIAH will be developed as a social marketplace for services. The paper presents the design decision we made and our proof-of-concept demonstrators and experiments.

1 Introduction

Digital research methods have recently started to enter the mainstream of humanities, arts and social sciences research. Digital arts and humanities have existed for years in specialised fields but the recent growth in the number of centres and research projects associated with digital methods in arts and humanities and social sciences indicate that we are at fundamental shift. But, digital arts and humanities is still a very young discipline where ad-hoc experiments dominate rather than systematic investigation. What is lacking is an infrastructure that would ensure that the state-of-the-art of these collaborations is preserved and integrated and that common best practices and methodological and technological standards are followed. The Digital Research Infrastructure for Arts and Humanities (DARIAH)[1] aims to be this infrastructure for Europe.

Until early 2011, DARIAH has been directly funded by the European Commission to prepare its organisational and technical framework. It is now funded by its member states and organisations, which currently include over 10 European countries. From 2012 onwards, DARIAH will move into production. By then, most of the national DARIAH projects will have started. DARIAH-EU will be organised into four virtual competency centres (VCCs) focussed on one particular area of expertise: (1) e-Infrastructure, (2) Research and Education Liaison, (3) Scholarly Content and (4) Advocacy. These VCCs bring together the national and topical humanities data centres, specialised research institutions and infrastructure service providers. The independence of these centres is paramount. Therefore, the DARIAH infrastructure has to be decentralised and lightweight.

This paper analyses the technical and scientific work in DARIAH's preparatory phase to set up DARIAH as such a light-weight, decentralised infrastructure. We aimed to prepare an infrastructure that on the one hand uses the innovations from the national initiatives and on the other hand ensures that the activities are embedded in the European DARIAH organisation. In the preparatory phase, we were looking for an infrastructure model that could express this. We wanted to avoid the impression of DARIAH as a 'network of roads', an image often used by developers to describe attempts to set up generic infrastructures. We, on the other hand, were looking for a model that would allow for the integration of services built around communities. To

[1] http://www.dariah.eu

978-0-7695-4597-4/11 $26.00 © 2011 IEEE
DOI 10.1109/eScience.2011.30

this end, we describe DARIAH as a social marketplace for services. This papers presents the results of our preparatory phase analysis how such a marketplace looks like.

The paper is organised as follows: In Section 2, we give a brief overview of how we tried to map the existing diversity in digital arts and humanities and in the DARIAH partner services. Section 3 develops our vision of an infrastructure as a marketplace of services. We present the overall architecture of the DARIAH network in Section 3.1. We demonstrate how we would like to develop the DARIAH partner sites as nodes in this network in Section 3.2, ensuring compliance of services with the DARIAH policies and mechanisms as outlined in Section 3.3. DARIAH will be presented to users in what we call service packages (Section 3.4). Finally, in Section 4 we discuss the experiments and demonstrators we developed as proof-of-concepts in the DARIAH preparatory phase.

2 Mapping Diversity

Diversity in the field of digital arts and humanities is larger than in other fields, as many disciplines and subdisciplines are included. From an infrastructure point of view, it is important to transfer this diversity into a collaborative information realm in order to pollinate exchange about these aspects for better ways of expressing them formally and continuously over time.

DARIAH aimed to capture this diversity and map it in two ways in its preparatory phase. Firstly, user requirements were mapped to 'scholarly primitives' (Unsworth), as it has been discussed in detail in [6] and [2]. These are generic actions by researchers and can be seen directly related to services an infrastructure has to fulfill. The assumption is here that research activities can be decomposed into the services that support these actions. We developed a mapping of the scholarly activities (primitives) towards services provided by DARIAH on the one hand and on the other hand to the requirements DARIAH has to outside service provides [2].

Our second mapping activity aimed to capture existing technologies in the DARIAH consortium to analyse gaps and ways of joining them up. To this end, we developed a research life cycle for digital arts and humanities and a mapping of existing DARIAH technologies onto it. We have found that DARIAH-related technologies can in fact cover a generic research lifecycle for arts and humanities. The full report can be downloaded from [9].

In this article, we want to concentrate on three key technologies that offer insights how existing tools and services at DARIAH partner sites can help enhance research processes in digital arts and humanities. They also show at the same time how diverse the technologies within the consortium are. The three projects are TextGrid, MIXED and eSciDoc. There are many others but we decided to only include production-ready solutions. These three projects also stand for developments of dedicated solutions to one specific problem (MIXED), for generic infrastructures that are used across disciplines (eSciDoc) and for the support of a specific task within digital arts and humanities (TextGrid).

TextGrid[2] is an example of a technology that supports core XML annotation in standard formats. It allows reusing and sharing TEI documents[3], an important community standard for deep scholarly annotations of text, as well as for their dissemination. As a complete solution for scholarly production it supports all parts of the research life cycle but its services can also be used for more specific tasks such as archiving research material on national in Grid infrastructures.

MIXED[4] complements TextGrid and has developed more sophisticated digital preservation services. It is a dedicated solution for archiving research material, developed by the DARIAH partner DANS. It uses a strategy of converting data to intermediate XML. If the datasets are later on to be disseminated again, they are converted from this generic format into a current vendor format of choice. The intermediate format can be easily adjusted to new format requirements.

The DARIAH partner Max Planck Digital Library co-develops eSciDoc[5] a generic e-Research platform to support research in Max Planck Society's institutes. Again, eSciDoc can be decomposed into disparate services that can be reused in other systems, for instance in the DARIAH archival software stack, as described in Section 4.4.

Both mapping activities have confirmed, that DARIAH cannot be a single monolithic infrastructure but that we need to apply an alternative approach. The analysis of user views and needs as well as the existing technologies led us to the development of architecture that would be able to accommodate such diversity. We found an inspiration in new concepts for social enterprise computing [1], which use Internet and social web technologies to create a social marketplace for services within an enterprise to connect with stakeholders and build products around communities.

DARIAH attempts to build services around communities, which can then be exchanged between communities in a virtual social marketplace that connects community workspaces with trusted DARIAH repositories of research data. This paper presents our attempts to plan and realise this aim.

[2] http://www.textgrid.de
[3] http://www.tei-c.org/index.xml
[4] https://sites.google.com/a/datanetworkservice.nl/mixed/
[5] http://www.escidoc.org/

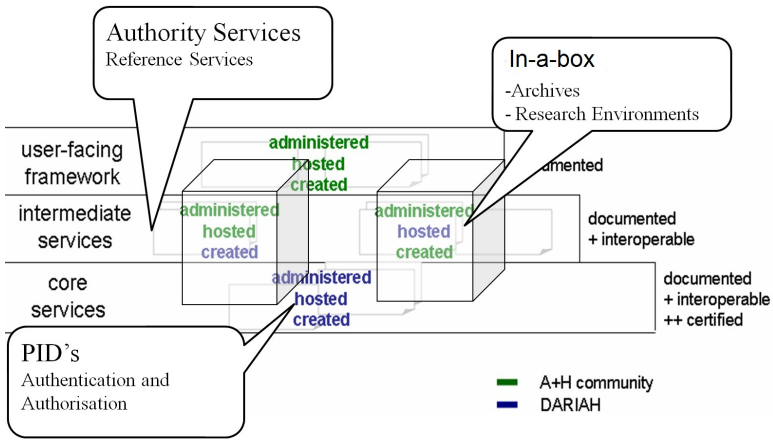

Figure 1. DARIAH architecture

3. Infrastructure as a virtual marketplace of services

DARIAH is designed to support the exchange of knowledge and services in dedicated virtual social marketplaces. A digital social marketplace is a framework to support advanced collaboration across diverse networks and specialised service providers. It has three main pillars [1]:

- Open APIs to expose reusable services (see Section 4.2)
- Composition and aggregation facilities to work with these services (see Section 3.4)
- Promotion of applications based on these services for several use cases. To this end, we have included dedicated education and outreach activities in DARIAH in VCC 2 and 4 as well as community demonstrators described in Section 4.

The digital marketplace is therefore a management layer for participants to expose and share services or create new ones between the communities they are part of. As a marketplace, it is an open environment where services can be promoted and exchanged. DARIAH sees itself as such a marketplace.

According to Figure 1, the DARIAH marketplace of services is organised around a loosely-coupled service-oriented architecture with three tiers: a core, an intermediate and a user-facing framework tier. Various services in each of these three tiers may interact in a single application. Core services are enablers of the marketplace. They are hosted and maintained by DARIAH to ensure their reliability, whereas higher-level services may be hosted and administered by other service providers. They are offered on the marketplace as parts of an open ecosystem of interacting services.

This section describes first the technical categorisation of the three tiers (development guidelines and shared infrastructure components), and subsequently reflects on their organisational and operational features.

3.1 Service markets

In Figure 1, each tier may open up different organisational contexts for managing service components. The full service catalogue of DARIAH can be found at [9]. The core technical infrastructure will be guaranteed by DARIAH as a platform for users to build their services upon. Its services are created, hosted and administered by DARIAH ensuring reliability and scalability.

The core layer includes light-weight services that serve to sustain the DARIAH infrastructure and establish coherent operation across the open DARIAH environment. This core layer will in the long term include a wide range of technical services. Immediately, we plan to expand our existing Persistent Identifier (PID) resolvers and an integrated community-based Authentication and Authorisation Infrastructure (AAI) for single sign-on, so that all DARIAH partners can benefit. These are essential for enabling interoperability across the heterogeneous data sources and decentralised services in the DARIAH ecosystem. The Persistent Identifier Service (PID) (see our experiment in Section 4.1) especially is a good example of how to enable citability of research objects and the openness of the DARIAH infrastructure. The DARIAH PID service links various system components with relevant policies. While there are numerous experiences on establishing PID services, DARIAH

faces the specific challenge of how to weave together diverse PID schemas that are currently in use in the DARIAH ecosystem [9].

The intermediate layer in Figure 1, the infrastructure service environment, has services that will be supported by DARIAH but not guaranteed. The national projects will collaborate with outside initiatives and researchers to build them. E.g., the DARIAH-DE supported Authority Mediation Service (AMS) deploys a network of reference data services, including library authority lists (e.g. Virtual International Authority File (VIAF)[6]) as well as various dictionaries, thesauri and gazetteers. As building these resources falls under digital scholarship, many DARIAH research partners are directly involved in setting them up. DARIAH-Serbia works on Serbian dictionaries while in Denmark they build DigDag, a digital atlas of the Danish administrative boundary units.

The user-facing framework (UFF) finally exemplifies another core principle of DARIAH. For the UFF, we document how to interact with the guaranteed core services but we also accommodate a collection of end-user tools contributed by research projects or third parties. Beyond mere documentation, tools and services ideally comply with the DARIAH service framework to foster interoperability with other DARIAH components (see Section 4.3).

Next, we describe ways of a light-weight partner integration into the DARIAH marketplace. The first one uses a reference architecture framework to harmonise technology development among the partner sites while the second deploys reference software solutions.

3.2 Developing the service stalls

A reference architecture is a proven way of helping to integrate disparate systems. It is part of the new social enterprise software solutions [11], because it attempts to develop collaboration around technological decisions and aims to provide guidance for the development of systems while maintaining their relative independence. Reference architectures are designed to mitigate change across multiple sites by multiple authors in different organisations. They help across domains and find new applications for the same or similar services. Global DARIAH applications can be customised towards local needs, and a reference architecture is one way to guide this localisation successfully.

The DARIAH reference architecture will provide best practices in system design and development. It will develop an architecture blueprint. In the preparatory phase of DARIAH, we experimented with a registry for the Logical View of a reference architecture, which analyses the functionalities. The Logical View leads directly to the DARIAH functional service descriptions of Section 3. The final registry for the reference architecture is currently under development, and DARIAH will lead the corresponding WP for DASISH ('Data Service Infrastructure for the Social Sciences and Humanities'), a new FP7-funded project to bring together all ESFRI social sciences and humanities projects.

Next to reference architectures dedicated software solutions help with participating in the DARIAH marketplace. Beyond services that correspond to the three conceptual layers of DARIAH, there will be services and software solutions that will make participation in DARIAH easier. In Figure 1, these are visualised as boxes cutting across the layers. Currently, we aim for two specific DARIAH-created solutions aimed at arts and humanities institutions wishing to create their own new digital research archives or digital research environments. Both 'In-a-box' solutions combine software that is installed and administered at the institution and 'connects' to the DARIAH central infrastructure services. In-a-box services are reference software packages that are created by DARIAH, yet hosted and administered by institutions. Their aim is to build capacities at arts and humanities research centres and accelerate uptake.

In Section 3.3, we now discuss how services can enter the DARIAH marketplace and how they can develop a more prominent place, before in Section 3.4, we analyse how services can be sold to the researchers by packaging them correctly.

3.3 Entering the marketplace: Compliance

In order to enter the DARIAH marketplace, services need to develop compliance, from being purely documented in the DARIAH registries towards being fully DARIAH certified. In an infrastructure that sees itself as a virtual social marketplace trust is one of the major problems. DARIAH will need to work as a provider of trust and our compliance work is supposed to ensure this.

Services (data and tools) shared by the arts and humanities community may have different integration levels, from the simple availability of documentation to services that are quality-certified by DARIAH-respective guidelines. In order to establish coherence across the horizontal tiers (core, intermediate and frontend), the DARIAH infrastructure defines integration frameworks for data and services. Compliance with these frameworks ensures that data and services can be contributed by decentralised parties. At the same time, they will be seamlessly accessible across the DARIAH research ecosystem. In other words, these frameworks ensure quality at three integration levels in the DARIAH architecture of participation. Services need to be at least documented, should be interoperable and can finally be fully DARIAH-certified.

At the lowest integration level, DARIAH expects compo-

[6] http://viaf.org/

nents in the DARIAH infrastructure to be well *documented* according to DARIAH standards. Moreover, documentation and transparency of research services may be required for good scholarly practice, and they may hence be the key to trust. One of the guiding principles of the DARIAH registries for services is that they have to be understood by humans first. DARIAH has started developing semantic registries to achieve this, which use social software collaboration environments. [10] discusses this approach for the TextGrid service registries.

The minimum requirement to foster *interoperability* across decentralised elements in the DARIAH ecosystem include mandatory metadata elements with respect to data object models, as well as guidelines for protocols and open APIs for services and tools. At the highest integration level, DARIAH aims to *certify* services based on existing best practices and relevant international standards. DARIAH partners are already active in standards for trustworthy data curation, which would affect both data sources as well as the objects contained in them. The certification will be light-weight reusing the insights from DANS, the Dutch DARIAH partner, and their Data Seal of Approval[7] while extending it towards general services.

Each compliance level is a step further into the DARIAH ecosystem, with more responsibilities but also more opportunities to benefit from — again following the general principles of a service marketplace in a social web. For example, only 'interoperable' services are capable of interacting with core infrastructure components, and only DARIAH-certified services offer reliable services, DARIAH commits itself to maintain. In the preparatory phase, we simulated the development of services into DARIAH-certified ones by enhancing the functionalities of existing ones, as described in Section 4. The general certification framework is currently under development but will generally be kept light-weight.

While the DARIAH compliance framework develops technological and infrastructure trust in the various DARIAH offerings, we have created service packages in order to deliver them. At first, we experimented with fully developed workflow-environments to aggregate services but found that such workflow-environments only provide partial solutions (e.g., for data flows). We have also made good experiences with workflow-environments for dedicated textual research services such as OCRing [7]. What is, however, generally needed, is a flexible means to combine general support services such as data curation services with compute services, where a researcher will immediately identify how this service package will help her research.

[7] http://www.datasealofapproval.org/

3.4 Packages from the marketplace

An infrastructure needs to provide a clear view to users on how services can be composed and aggregated. To this end, DARIAH has created so-called service packages that support various user needs. These packages address particular requirements of dedicated user communities. A package can consist of a mix of both support (for instance, advice on preservation formats or legal issues) and technical services, with restrictions enforced by interoperability between the technical services. A package can be defined by DARIAH, for example a service bundle to build local capacity, or by any community of users, for example platforms to support history research. A package may be more generic and therefore of interest to more than one community of users, for example a package of various text-mining tools.

In Figure 2, we can see how a researcher can make use of more than one package, and moreover how a package can be composed of a number of services. A front-end technical service utilises intermediate and core services to provide the application functionality required. However, it should be possible for a package to independently use a core or intermediate service without the mediation of a front-end service.

User roles, packages and component tools and services are shown in Figure 2. Users of the support and technical services are not just scholars. For instance, an arts and humanities organisation, wishing to establish a new digital archive, may use the DARIAH Building Capacity package for the technical infrastructure required and the DARIAH Data Curation package to develop the institutional know-how. The History Tools package in Figure 2, used by a history scholar, may not be created by DARIAH, but by a group of historians interested in text analysis. Furthermore text-mining tools may be added to the DARIAH ecosystem by a tools developer, but this front-end service still utilises intermediate and core services developed by DARIAH.

In Figure 2, our mapping work (see Section 2) helped us define DARIAH packages, while the service catalogue, which we discuss in the next Section 4, is based on bringing together the support services the existing DARIAH partners are already running and on developing the technical services using proof-of-concept experiments. In the DARIAH preparatory phase, we have spent a substantial effort on simulating the future DARIAH marketplace with community demonstrators and technical experiments. Our interest was to demonstrate that development can take place independently without too much of a compromise in terms of consistency. We considered this work also necessary to communicate our technical work as something that directly serves community needs. In the next section, we present some of these experiments and demonstrators as well as national DARIAH member projects for the preparatory phase,

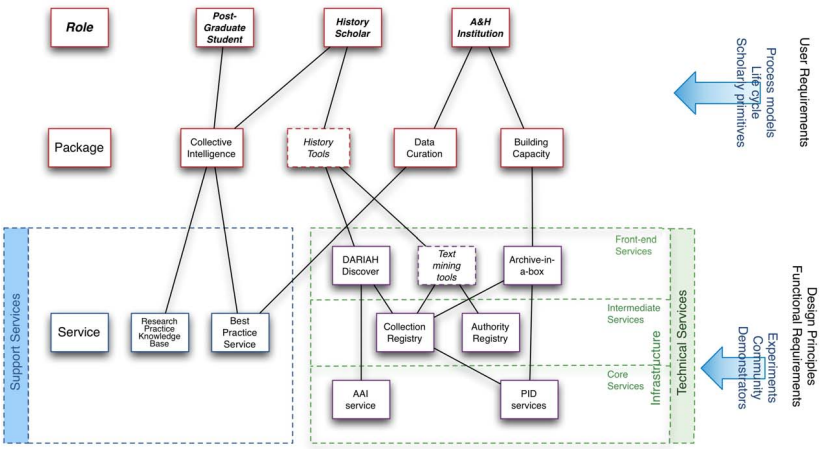

Figure 2. DARIAH-packages

which we consider to be building blocks of the DARIAH production infrastructure.

4. Simulating the marketplace: THE DARIAH Service Catalogue and Proof-of-concept

In the DARIAH preparatory phase, the technical proof-of-concept work was divided into at least two principal activities. The first one investigated how to enable the DARIAH infrastructure as a social marketplace and validated the underlying concepts in a series of experiments. This set of experiments is based on the mapping of primitives to infrastructure functions and therefore on how to build services around communities. The second principal technical activity were demonstrators to use existing DARIAH technologies to support communities, which have historically been at the forefront of developing digital humanities: archaeology and classics as well as textual studies. Details of the demonstrator work can be found in [2]. In this paper, we would like to concentrate on discussing examples for each layer of the infrastructure in Figure 1 and how they help enable the marketplace of services. For each example service, we discuss the envisioned functionalities and our proof-of-concept work.

First, we discuss the PID service that will be a core component of the DARIAH certified and guaranteed services. It is a good example for a core service not just because it is generally needed in any online federation of data resources but also because it can be considered to be a stable service based on years of experience at the DARIAH partners.

4.1 Core Infrastructure — Persistent Identifier Services (PID)

The DARIAH PID service provides the user (institution or researcher) with persistent identifiers for their digital research objects in the DARIAH ecosystem. It scales current partner PID services to a European level. DARIAH recognised early on that the use of PIDs within the new infrastructure is imperative. When a researcher cites an article or dataset in her (hardcopy) thesis, she needs to be assured that the citation itself will always lead to the original resource she has used. Moreover, creating relationships between resources, such as information about a researcher and the articles she has published, requires a permanent mechanism in which the tie between the different resources can persist.

Many DARIAH members already provide their own PID solutions.[8] The PID experiments were especially designed to accommodate these local solutions and to be able to use PIDs in the heterogeneous environment of existing archives, DARIAH in-a-box services, etc. To this end, a prototype PID-system has been implemented that:

- allowed clients to access the physical location of a resource, based on the HTTP protocol, regardless of the PID's origin. The DARIAH pluggable PID meta-resolver can handle any kind of existing PID (Handle, DOI, ARK, etc.).
- enabled users to refer to parts of resources (e.g. a specific 10 seconds of a video clip).
- enabled users to refer to particular representations of resources, for instance the HTML landing page.

[8] E.g.: http://www.dans.knaw.nl/content/categorieen/diensten/persistent-identifier

During design and development of the experiment, it became clear that a number of the requirements, such as content-negotiation and trust feedback, need to be the responsibility of the content-provider disseminating the resources to the client, instead of the responsibility of the PID service. What this means for DARIAH is that simply choosing a PID framework or system is not enough to ensure that resources can be referenced freely and persistently. Thus, we decided to enhance the archive-in-a-box solution with ready-made solutions that adhere to the standards that DARIAH needs to set for content-negotiation, trust feedback and part-addressing (see also Section 4.4). However, organisations that already have software in place for resource discovery and delivery need to analyse their systems and implement additional functionality or wrappers to be able to fully profit from the enhancements DARIAH will bring to the data infrastructure. We have started to develop a corresponding reference architecture to help with technical and administrative decisions.

Next we discuss our attempts to experiment with open interfaces to resources as the decisive step to integrate outside services into the DARIAH marketplace for the intermediate layer in Figure 1.

4.2 Intermediate Infrastructure — Open APIs

A particular emphasis was placed in the DARIAH preparatory phase on the development of open APIs that enable seamless communications between DARIAH service providers. A good example is our OAI-ORE[9] experiment. Based on the frameworks developed in [3] and the idea to use the Open Archive Initiative standard OAI-ORE together with ATOM feeds for the exchange of information, we aimed to simulate a federation of content repositories. Details of the architecture framework are discussed in [4]. Our prototype linked the Grid-based TextGrid repository with a Fedora repository, which catered for data analysis (i.e. XQuery capabilities on XML/TEI objects) across repositories, and other conceivable applications. Compared to existing solutions for the federation of repositories such as CMIS [8], OAI-ORE has the advantage that it is an open standard and is already used in many Linked Data applications, for instance for the Europeana Linked Data cloud.[10] Technically speaking, the experiment has shown that such a federation of repositories based on OAI-ORE and ATOM is feasible but it has also shown that much more work needs to be done to agree on a possible interchange format that provides deep enough information for research. Commonly used examples such as Dublin Core have not convinced, as they lack the necessary detail for research.

Next to the OAI-ORE experiment, we ran several experiments to open up existing systems. We tried to enhance the UK arts-humanities.net platform, which contains a human-readable knowledge base of tools and methods in the digital arts and humanities. We added machine-readable service descriptions and linked these to the existing TextGrid service registry. The descriptions of digital methods contained in arts-humanities.net have shown to be a good interface to browse the collections.

As detailed in [5], we also attempted to build a simplified web-based interface to European Grid storage. We re-engineered the REST API of the Amazon S3 storage service as an interface between the grid environment and the repository. However, it turned out that Amazon does not allow to re-engineer S3 so we turned our attention to OpenNebula[11] to provide access to distributed data infrastructures. We have a working installation in London and at DARIAH-DE for the archive-in-a-box and plan to link in further resources soon.

Finally, one of our community demonstrators worked on transforming an existing legacy application into a service-oriented architecture. The DARIAH consortium agreed to migrate the legacy application 'Archaeological Records of Europe'- Networked Access (ARENA) at the Archaeology Data Service (ADS) partner into a service-oriented architecture. The existing metadata search portal service was enhanced by using DARIAH web services to expose the attached databases as autonomous services. The local search services publish themselves to a service registry where they can be accessed by a client. In the case of ARENA, the services are either compliant search services for archeological resources or 'wrapped' services based on legacy protocols such as Z39.50.[12] With the ARENA demonstrator, DARIAH aims to show that open search services can be one way of integrating the many heterogenous data resources in the arts and humanities.

Next we discuss the user-facing framework (UFF). As the UFF shall integrate outside applications, we concentrate not on a DARIAH preparatory phase experiment but national projects and how these interact with the DARIAH infrastructure. The French ISIDORE platform was separately funded from DARIAH but will play a key role in the construction phase as part of the French contribution.

4.3 User-Facing Framework: DARIAH-Discover

The UFF will contain DARIAH-Discover with a number of search and browsing services across all the collections. The user will have the opportunity to browse a collection or perform a textual search across one or all the collections. As analysed in [6], browsing even more than searching is

[9] http://www.openarchives.org/ore/
[10] http://data.europeana.eu/
[11] http://www.opennebula.org
[12] http://www.loc.gov/z3950/agency/

a key scholarly activity in the humanities. The DARIAH demonstrator ARENA2 client, for instance, offers a facetted browsing service according to the core facets for Humanities research: what, where and when.

Next to ARENA, a good example for DARIAH-Discover is the French ISIDORE research platform,[13] which is a Linked Data application to provide a search interface to a wide range of research data sources in the arts and humanities and social sciences. ISIDORE harvests metadata as well as full records from French research data resources. Once harvested, the information is transformed into RDF, stored in a triple store and enriched with outside references to vocabularies, thesauri, etc. ISIDORE is thus a web of data application, and DARIAH will reuse some its services and scale them to other European countries and languages.

In the next section, we discuss how we bring all the architecture layers together in packaged solutions.

4.4 DARIAH Archive-in-a-box

The archive-in-a-box service will provide an institution with the facility to install software on its servers in order to create a digital asset management system for its research community. It is aimed at making the participation in the DARIAH ecosystem as seamless as possible and offers services to support the full lifecycle of digital research objects. In particular, the service attempts to make the management of research objects as simple as possible. For instance, technical metadata will be automatically generated to which a user may add additional descriptive metadata.

The service consists of a complete technological solution for an institution (or organisation) wishing to create a digital humanities archive. The archive-in-a-box complies with all the DARIAH service requirements and shall provide an interface by which collections that are registered with DARIAH are made available. An extension to this service is the digital preservation service, which provides tools to access grid and cloud storage solutions. This set of storage interfaces can be also installed independently of the rest of the archive-in-a-box service.

Most of the core systems of DARIAH partners already provide solutions for the digital archives. The MaxPlanck eSciDoc system is the most advanced one in many ways. That's why we decided to build a demonstrator in the preparatory phase that showcases its possibilities using digital humanities research objects. The purpose of the DARIAH TEI Demonstrator was to make it easy for humanities researchers to share TEI-encoded texts with others inside and outside their institutions, and to compare their encoding practice within the TEI community. We installed, for instance, central schema instances.

[13] http://www.rechercheisidore.fr

5. Conclusions

This paper has presented the infrastructure vision for DARIAH according to the technical and scientific work in the preparatory phase. DARIAH will not be one technical solution, but many, according to community activities and willingness to collaborate. Key to the success of such an infrastructure is to understand clearly how users will interact with multiple technical solutions. Users will see services built around their communities while their communities can exchange these services at virtual marketplaces. This paper has shown, how DARIAH addresses the large diversity in the field of digital arts and humanities with this architecture. In the preparatory phase we have furthermore shown that technical solutions exists to address this diversity, using demonstrators and experiments.

References

[1] K. Al-Begain, C. Balakrishna, L. A. Galindo, and D. Moro. *IMS: A Development and Deployment Perspective*. Wiley, 2009.

[2] S. Anderson, T. Blanke, and S. Dunn. Methodological commons: arts and humanities e-science fundamentals. *Phil. Trans. R. Soc. A*, 368(1925), August 2010.

[3] A. Aschenbrenner, T. Blanke, D. Flanders, M. Hedges, and B. O'Steen. The Future of Repositories? Patterns for (Cross-)Repository Architectures. *D-Lib Magazine*, 14(11/12), 2008.

[4] A. Aschenbrenner, T. Blanke, M. W. Küster, and W. Pempe. Towards an open repository environment. *J. Digit. Inf.*, 11(1), 2010.

[5] A. Aschenbrenner and S. D. Flavia Donno. Infrastructure for interactivity — decoupled systems on the loose. In *Proceeding of the Third IEEE International Conference on Digital Ecosystems and Technologies*, 2009.

[6] A. Benardou, P. Constantopoulos, C. Dallas, and D. Gavrilis. Understanding the information requirements of arts and humanities scholarship. *International Journal of Digital Curation*, 5(1), 2010.

[7] M. Bryant, T. Blanke, M. Hedges, and R. Palmer. Open source historical ocr in the ocropodium project. In *Proceedings of European Conference Research and Advanced Technology for Digital Libraries (ECDL)*, 2010.

[8] CMIS. Content mangement interoperability services. http://goo.gl/ur4NY, Accessed 01/07/2011.

[9] DARIAH. Dariah — technical report. http://goo.gl/AUuNF, Accessed 01/07/2011.

[10] M. W. Kuester and C. Ludwig. Digital ecosystems of ehumanities resources and services. In *Proceeding of the Second IEEE International Conference on Digital Ecosystems and Technologies*, 2008.

[11] G. Mueller. A reference architecture primer. http://goo.gl/74BAe, Accessed 01/07/2011.